GIS Data Layer Design and Creation Guidelines

Last revision: February 24, 2006
Effective date: March 1, 2006

Purpose

This document identifies steps that need to be followed for creation of all Geographic Information Systems (GIS) data sets related to projects of the North Coast and Cascades Network (NCCN) Inventory and Monitoring (I&M) Program. It is envisioned that these guidelines will also be followed in the creation of non-I&M GIS datasets to improve accuracy and usability of all NCCN GIS data.

These guidelines will:

- identify the process by which spatial products are to be developed and maintained by the NCCN GIS
- identify Quality Assurance (QA) procedures and data set acceptance criteria to be followed when generating and maintaining spatial products

Scope and Applicability

These guidelines should be used by NCCN GIS staff, Project Leads, Data Mangers, cooperators, and contractors for creation of spatial data relating to NCCN I&M projects. Specifically, all data that will be maintained in one of the NCCN Digital Libraries or submitted to one or more of the NPS national data clearinghouses will be subject to these guidelines. All parties creating and/or submitting GIS-related data to NCCN should work with NCCN GIS staff in all stages of data creation, along with using guidance within this and other documents listed in Guidelines and SOPs section of this document.

Note: NCCN uses ESRI® GIS products, such as ArcView® 3.x, ArcInfo®, and ArcGIS®. NCCN has not yet transitioned to the ESRI geodatabase format due to resource constraints, including training and development time, and the cost of licensing for enterprise-level database administration. This option also does not fit as easily into current work flow practices by network staff. Although NCCN uses ESRI products, this should not be taken as our endorsement of these products.

Definitions and Acronyms

Attribute table	A tabular file containing rows and columns. Attribute tables are normally associated with a class of geographic features. Each row represents a geographic feature. Each column represents one attribute of a feature.
Cardinality	Term describing how many features in one spatial data set are related to how many records in the attribute table of another data set (i.e., one-to-many, many-to-many, etc.).
Coverage	The ESRI® ArcInfo® spatial vector data model. It generally represents a single set of topologically related geographic objects such as roads, parcels, soil units or forest stands in a given area. These features are symbolized by points, lines, polygons, routes or regions. A coverage supports the georelational model - it contains both the spatial (location) and attribute (descriptive) data for geographic features.
Database	A collection of data organized according to a conceptual structure describing the characteristics of the data and the relationships among

	their corresponding entities. For example, a GIS database includes data about the position and characteristics of geographical features.
Data dictionary	Documentation of a data set that defines attributes such as domains, range limits, and validation rules.
Directionality	Term describing the direction of a relationship between two tables (i.e., source or destination).
Domain	The realm of possible values for a data field. The domain may be constrained either fully or in part by a pick list, range limits and validation rules.
ESRI®	Environmental Systems Research Institute, Inc. of Redlands, CA. Makers of ArcView, ArcInfo and ArcGIS software.
FGDC	The Federal Geographic Data Committee is an interagency committee who develop policies, standards, and procedures for organizations to cooperatively produce and share geographic data.
Foreign key	A relationship or link between two tables which ensures that data stored in a database are consistent. The foreign key link is set up by matching a column(s) in one table to the primary key column(s) in another table.
Geodatabase	The ESRI name for 'geographic database.' The Geodatabase model is an ArcGIS version 8.0 and above data format. A geodatabase represents geographic features and attributes as objects and is hosted inside a relational database management system (DBMS).
GIS	A Geographic Information System (GIS) is a computer system for capturing, storing, checking, integrating, manipulating, analyzing and displaying data related to positions on the Earth's surface.
GPS	Global Positioning System
I&M	Inventory & Monitoring Program of the National Park Service.
Line	A type of vector geometry having length and direction but no area, connecting at least two x,y coordinates. Lines represent geographic features too narrow to be displayed as an area at a given scale, such as contours, street centerlines, or streams, or features with no area that form the boundaries of polygons, such as state and county boundary lines. Lines, or segments of lines, are referred as "arcs" in ESRI software.
Logical consistency	An evaluation of the interaction between the values of two or more functionally related attributes. If the values of one attribute change, the values of functionally related attributes must also change.
Metadata	Data about the data set. Usually provided in the form of a text, xml or html document with information on the data set's quality, coordinate system and projection, attributes, distribution and citation. In the National Park Service, it is generally a file compliant to the FGDC Content Standard for Digital Geospatial Metadata and NPS Metadata requirements.
NCCN	North Coast and Cascades Network: *http://www1.nature.nps.gov/im/units/nccn/*
NPS	National Park Service
Physical consistency	An assessment of the topological correctness and geographic extent of the GIS database.
Point	A type of vector geometry used to represent point features. A point is defined by a single x,y coordinate pair.
Polygon	A type of vector geometry used to represent polygon features. A polygon is defined by one or more rings, with a ring defined as a path that starts and ends at the same point.

Positional accuracy	A measure of how well each feature's spatial position in the data set matches reality.
Primary key	A unique identifier for each record in a relational table. It can either be a user-defined attribute that is guaranteed to be unique or it can be generated by a database management system, such as the globally unique identifier (GUID) created in Microsoft® SQL. Primary keys may consist of a single attribute or multiple attributes in combination.
Random errors	Random errors result from unknown or accidental causes such as improper recording of spatial or attribute information.
Raster	A spatial data model that defines space as an array of equally sized cells arranged in rows and columns. Raster coordinates are contained in the header of the array. The attribute value of each cell represents the value of the feature. Groups of cells that share the same value represent the same type of geographic feature.
Referential integrity	An assessment of the association of related tables based upon their primary and foreign key relationships. Ensures that the primary and foreign keys exist and that all foreign key values entered match an existing primary key value.
RMS error	Root Mean Squared Error. Registration errors are introduced when paper maps, film separates or digital images are registered to a digitizing board or to images with known coordinate locations. The RMS value represents the amount of error between original and new coordinate locations and is calculated during the digitizing or transformation process.
Shapefile	An ESRI GIS data format that stores non-topological geometry (see definition of coverage) and attribute information for the spatial features. The geometry for a feature is stored as a shape comprised of a set of vector coordinates. Shapefiles can support point, line and area features.
SOP	Standard Operating Procedure.
Systematic errors	Systematic errors are those errors that follow some fixed pattern and are introduced by data collection procedures and systems.
Topology	The spatial relationships between connecting or adjacent coverage features (e.g., arcs, nodes, polygons, or points). For example, the topology of an arc includes its from- and to- nodes, and its left and right polygons.
Validity	A means of enforcing attribute accuracy of the database through a set of rules that control data entry.
Vector	A coordinate-based data model that represents geographic features as points, lines or polygons.

Overview

This document examines steps involved in design, creation, archival and maintenance of a NCCN I&M data set, as well as the management and planning issues associated with each of the above steps. The quality and usefulness of a final data set depend, almost entirely, on how well the data set was designed. Good data set design allows the data set to be viewed in its entirety so that interaction between elements can be evaluated and permits identification of potential problems and design alternatives. Without a good data set design, there may be irrelevant data that will not be used, omitted data, no update potential, inappropriate representation of entities, lack of integration between various parts of the data set, unsupported applications and/or major additional costs to revise the data set. Figure 1 summarizes the major steps in date set design and creation, with emphasis given to the data set design process.

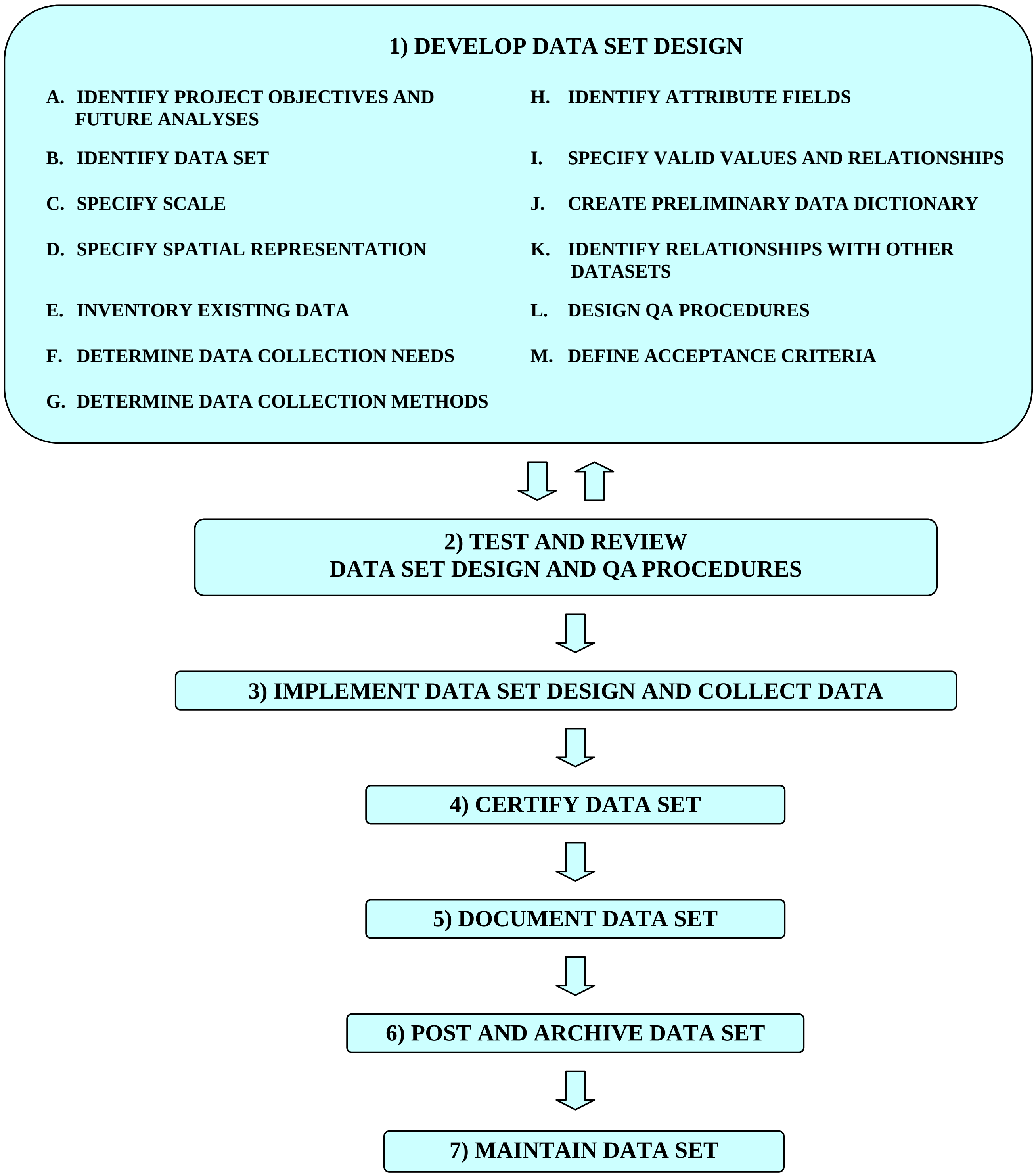

Figure 1. Steps in designing and maintaining an I&M-related spatial data set.

General Procedures

For general guidelines on NCCN project planning and data acquisition, processing, posting, maintenance, and archival, consult the Data Management Plan (Boetsch et al. 2005). More specific information on non-GIS data can also be found in a number of guidance documents posted on the NCCN public website (http://www1.nature.nps.gov/im/units/nccn/dm_docs.htm). The majority of this document has been adapted from documents listed in **Credits** and **Reference Documents** sections at the end of this document.

1. Develop Data Set Design

   A. Identify Project Objectives And Anticipated Analyses
   The Project Lead or Principal Investigator will provide the GIS Specialist with a project proposal stating project hypotheses, objectives, and study design. The Project Lead or Principal Investigator and GIS Specialist will discuss how the created data sets will be used with existing NCCN data. Appendix 1 lists and describes spatial analysis functions that can be performed on GIS data sets to maximize the utility of spatial information.

   B. Identify Data Set
   Depending on study design, the Project Lead or Principal Investigator, with the help of the GIS Specialist, will determine which spatial data set(s) will be needed for the project. Several data sets might be needed for the same project. For example, an inventory of forest carnivores might need one data set representing potential carnivore habitat, a second one representing a sampling frame (blocks), and a third representing potential locations for camera stations within chosen blocks. In addition, a fourth data set might be created when the coordinates of the actual camera station locations are recorded in the field.

   C. Specify Scale
   The Project Lead or Principal Investigator needs to know the project's accuracy requirements to determine the appropriate scale of future data sets. Accuracy requirements should be based on the size of the smallest feature to be mapped and the resolution or scale of other data sets used in conjunction with field data collection. National Map Accuracy Standards (US Bureau of the Budget 1947) dictate the accuracy (in feet or meters) on the ground for common scales. Consult NPS GIS Data Specifications for Resource Mapping, Inventories and Studies (National Park Service GIS Program 2005) and GIS Product Specifications (NCCN 2005b) for NPS and NCCN I&M project accuracy requirements. Required accuracy will determine the quality of input necessary and the amount of data that may be created. For example, projects that do not necessitate high accuracy might save money and time by using coarse scanning or digitizing methods or by using satellite image interpretation rather than very careful and detailed digitizing or field data collection.

   D. Specify Spatial Representation
   The Project Leads or Principal Investigators, with the help of GIS Specialists, need to determine whether the data set will be best represented by points, lines, polygons or rasters. For example, a telephone pole may be represented as a point feature, a road as a line feature, and wetland as an area feature. Spatial representation should be based on types of manipulations that might be undertaken. For example, a building might be represented as a point if building density in an area needs to be calculated. A polygon representation is best, if one is trying to determine the amount of impervious surface in a watershed. A raster might

be a good alternative to area features if certain kinds of spatial analysis, such as wildlife habitat modeling or watershed analysis, will need to be performed on the data set.

Determination of spatial representation will also depend, in part, on how the data set will be designed and managed. Vector data sets are easily managed within a GIS or relational database, whereas raster data sets are more appropriately managed through GIS or remote sensing software. Refer to the Spatial Data Integration (NCCN 2006g, in development) and Design Standards for Project Databases (NCCN 2006b, in development) documents for more guidance on project database design.

E. Inventory Existing Data
Once the needed data sets are identified, the Project Lead or Principal Investigator must coordinate the development of each data set with the GIS Specialist. The GIS Specialist will inventory Park, Network and other agency GIS databases to identify any existing data sets, including relevant Global Positioning System (GPS) data, similar GIS data available at different scales, incomplete GIS data, and other data not currently in a digital format. Appendix 2 provides a list of criteria for which each potential source for a GIS data set should be evaluated. All of the listed criteria need to be documented for each potential data source. If a particular data source is then used to build a GIS data set, some of this information will become part of the permanent metadata.

F. Determine Data Collection Needs
If needed GIS data do not exist or cannot be derived from existing GIS data, the Project Lead or Principal Investigator will determine what spatial information will need to be collected. S/he will work with the GIS Specialist to make sure the collected data will be of appropriate spatial scale and representation (determined in Sections C and D) and compatible with current Park or Network GIS data.

G. Determine Data Collection Methods
There are many forms of GIS data collection, which are described in Appendix 3. Gathering field data and management of data in a GIS is estimated to be 80% of a GIS's cost (Trimble 2000). The Project Lead's or Principal Investigator's selection of data collection methods should depend on project scale and spatial representation requirements (determined in Sections C and D), Park or Network GIS and GPS capabilities, time requirements for each collection method, and amount of funding available for the data collection and management stages of the project.

H. Identify Attribute Fields
A diverse group of data users should be invited to collaborate on determining the set of attributes collected for each feature. To ensure compatibility among data sets and to meet Park and Network needs, collect standard attributes whenever feasible. Consult Design Standards for Project Databases (NCCN 2006b, in development) and GIS Naming Conventions (NCCN 2005a) for NCCN-required core tables and attributes and their domains. Determine for each additional attribute the following:

i. Attribute data type - Are the attributes recorded as numeric, date, time or character string?

ii. Input restraints - Will the collector need to pick from a list of choices provided in a menu or will it be free text? Should the values be required to be filled in or will the information be considered non-critical? Should the values be required to fall within a predefined range?

iii. Measurements - What are the units of measurement?

If the data set will be merged or combined with data from a different source, make sure the attributes are identical (i.e., the same data type (numeric, character string, etc.) and width) and add an additional attribute to keep track of the source of each feature. In addition, data values should be consistent among data sets (e.g. “Sol Duc River” should be spelled the same way in each instance it appears. Query functions tend to be case-sensitive in GIS software, so values must also have consistent cases.

I. Specify Valid Attribute Values And Relationships
The Project Lead or Principal Investigator, with the help of GIS Specialist and/or Data Manager is responsible for determining the range or set of values each attribute field within the data set is allowed to have. In addition, they are responsible for determining any relationships an attribute field may have to another field and the nature of that relationship.

A value is a measurable quantity which a measurement may take that is either assigned or determined by calculation. For each attribute, the Project Lead or Principal Investigator determines the range, pick-list, or type of values the users can enter by setting validation rules. Examples of validation rules are: DBH_inch values must be between 0 and 100, Life_stage values must be ‘adult’ or ‘juvenile.’

Another type of validation rule might involve a relationship between two or more fields. Values in an attribute field within a data set often determine values in other attribute fields. For example, if a feature in a landcover classification has a value of “water” in the attribute “Cover_type” then the value of attribute “Slope” is predetermined to be “0,” and if it is not recorded as such, represents a logical consistency error.

J. Create A Preliminary Data Dictionary
After inventorying current GIS data sources and determining data collection needs for the project, the Project Lead or Principal Investigator should document the feature and attribute information to be collected in the field or from other existing sources. In this context, "preliminary data dictionary" refers to creating a list of features and their associated definitions and attributes to be collected. Below are the steps to preliminary data dictionary design and implementation. See Section 5 of this document and Design Standard for Project Databases (NCCN 2006b, in development) for further guidance on documenting a project data set.

i. Make a list of all the real world physical entities to be mapped.

ii. Categorize the list into point, line and polygon features.

iii. For each feature, identify the attributes that will be recorded and their data type, input constraints, measurement units, and text description.

iv. For each attribute, identify relationships to other attributes.

At this stage, the GIS Specialist is responsible for creating a GPS data dictionary if GPS data will be collected for the project. The GPS data dictionary should match the design of the project database to facilitate data import. As many attributes as possible should be entered in the field using the GPS Data Dictionary to reduce data entry time in the office and to minimize the errors that can be introduced when data are transferred from field data sheets into the project database. Field data sheets are still recommended as a backup in case of GPS failure or electronic data loss.

K. Identify Relationships With Other Data Sets (spatial or tabular)
The Project Lead or Principal Investigator, with the help of GIS Specialist and/or Data Manager, must specify how the data sets are related with other data sets, so these relationships can be created in a GIS project or a relational database.

A relationship is an association between two objects. These objects can be non-spatial (tables) or spatial (features). For example, a parcel of land (feature) may be associated with an owner (table) or a land-use zone (feature). With a relationship, one can define which column in a feature's attribute table (primary key) and which column in another table (foreign key) share the same values. In this example, the parcel of land would contain a parcel id (primary key), which would relate to a parcel id in the owner table (foreign key), and a land-use zone id (foreign key), which will relate to the zone id in the land-use zone spatial data set (primary key).

One object can participate in many relationships. Relationships are described based on their directionality and cardinality. See Spatial Data Integration (NCCN 2006g, in development) for a detailed description of different types of relationships and instructions on how to establish relationships in GIS.

L. Design Quality Assurance (QA) Procedures
The GIS Specialist is responsible for determining which QA procedures should be followed throughout the GIS data set creation. The selection of QA procedures will depend on the project data collection methods. The following categories of QA should be addressed, if applicable, for each data set and associated tables:

i. Completeness - an indicator of whether each feature or entity is present in the data set, and whether all of its attributes are present. The completeness measured by the provider is a relative measure, comparing the data set's objects versus what it is intended to represent. Completeness for the user, in assessing the data set's fitness for use, must support consideration of not only the provider's completeness measure, but also whether the represented set of features or entities is compatible with the user's application requirements. See GIS Product Specifications (NCCN 2005b), Design Standard for Project Databases (NCCN 2006b, in development) and GIS Naming Conventions (NCCN 2005a) for NCCN I&M data set standards.

ii. Attribute Accuracy - describes how well the assigned attribute values match the actual characteristics of the objects.

iii. Logical consistency - describes the structural integrity of a data set. This is concerned with assurance that identified constraints on data keys, attribute domains, and key and attribute interrelationships are observed. If the value of one attribute changes, the values of functionally related attributes must also change. For example, in a database in which the attribute SLOPE and the attribute LANDUSE are related, if LANDUSE value is "water," then SLOPE must be 0, as any other value for SLOPE would be illogical.

iv. Physical consistency - the topological correctness and geographic extent of the database. Physical constraints applicable to point data are issues of location: do neighboring points violate minimum distance requirements? Tests to verify this constraint can be automated with the GIS software; and violations of this type of error may be related to issues of positional accuracy.
Physical constraints applicable to line and polygon data consider what provides a complete and accurate indication of each object and how the objects relate to each other:
1) Are all objects completely described graphically?
2) Do any objects contain overshoots or undershoots?
3) Do objects intersect only where intended?
4) Do any objects exist twice?
5) Are any objects too close?
6) Are any polygons too small (sliver)?
7) Do any polygons overlap?

v. Referential integrity - the association of related tables based upon their primary and foreign key relationships. All tables must have primary and foreign keys and associated sets of data according to predefined rules for each table.

vi. Positional accuracy - measure of how well each spatial object's position in the database matches reality. For example, positional errors can result from incorrect cartographic interpretation, if line segments were digitized with fewer vertices than necessary, or from digital storage precision inadequacies. These errors can be random, systematic, and/or cumulative in nature. Positional accuracy must always be qualified because the map is a representation of reality.

1) Random errors result from unknown or accidental causes such as improper recording of spatial or attribute information. Random errors will always be a part of any data, regardless of form. Random error can be reduced by tight controls and automated procedures for data entry and by checking data automatically and visually at various stages in the processing cycle.

2) Systematic errors are those errors that follow some fixed pattern and are introduced by data collection procedures and systems. Systematic errors must be removed when data are converted to a different format. Implementation of visual and automated review of data in the beginning of the conversion process will help detect systematic errors.

Quality Assurance/ Quality Control Procedures for ITAM GIS Databases (Johnston et al., 1999) describes specific methods for assessment of each of these QA categories.

M. Define Acceptance Criteria
Defining acceptance criteria is probably one of the most troubling parts of the GIS project. It requires knowledge of the data model and database design, as well as the user needs and application requirements, and depends on project schedule, budget, and human resources. Define acceptance criteria for the data set by answering the following questions:

i. Which errors are acceptable?

ii. Are certain errors weighted differently than others?
Each attribute should be reviewed to determine if it is a critical attribute and then weighted accordingly. Primary and foreign keys should not contain any errors, since errors in these attributes will result in errors being propagated throughout the database.

iii. What percentage of error constitutes a rejection of data?
If a GIS data set has 10 features and each feature contains 10 attributes, what is the percentage of error if one attribute is incorrect? Is it 1 percent or 10 percent? If one subscribes to the theory that all of the attributes make a feature correct, then the entire feature is in error, making the error rate 10 percent. On the other hand, if only one attribute is incorrect for a feature, and it is treated as a minor error, then the error rate is 1 percent because one out of a possible 100 attributes is incorrect. Additionally, the cartographic aspect of data acceptance should be considered. A feature's position, rotation, and scaling must be taken into account when calculating the percentage of error, not just its existence or absence. Consult GIS Product Specifications (NCCN 2005b) for information on NCCN data set accuracy requirements.

iv. Error detection
Once the acceptable percentage of error and the weighting scheme are determined, methods of error detection for data acceptance should be established. These methods are the same as those employed during the data conversion phase (see Sections 1L and 3). Checkplots (selected cells of a grid overlaying the data set) are compared to original sources and automated database checking tools are applied to the data. Very large databases may require random sampling for error detection.

2. Test And Review Data Set Design

The Project Lead or Principal Investigator is responsible for completing a trial run of data collection with the project field crew leaders before commencing actual data collection. This will help everyone understand the features and attributes to be collected as well as ensuring familiarity with the equipment and technology. The Project Lead or Principal Investigator can expect to revise the attributes and sometimes the spatial representation of features as the data set design is tested. Below are the steps involved in testing the data set design:

A. Provide necessary training in use of GPS or GIS equipment involved in the data collection stage of the project. GIS Specialist and/or Database Manager must be closely involved in this step.

B. Test the data set design under field conditions, if applicable, by collecting trial data. Take field crew to the project site and start mapping, or run through a digitizing, scanning or classification exercise.

C. Check for other features that should be in the data set. Decide if additional data can be collected with minimal effort. For instance, if the project was to map an interpretive hiking trail, it will cost little additional effort to map the signs as one maps the trail.

D. Evaluate the attribute information collected during the trial run and make sure no necessary attributes were left out.

E. Implement trial QA procedures.

F. Use the trial data set to create relationships with other data sets or tables outlined in data set design and test their functionality.

G. Make and document any necessary changes to the data set design, data and metadata collection protocols, data entry protocols, or QA procedures, or acceptance criteria.

3. Implement Data Set Design And Collect Data
The Project Lead or Principal Investigator is responsible for collecting necessary data for the project. Depending on the data collection method, refer to one of the three sections below for instructions concerning data collection. Regardless of data collection method, field crews are required to QA their data on a systematic basis through the field season. For example, the Project Lead or Principal Investigator might designate one day after each week in the field for entering, processing, and/or checking the accuracy of data.

A. GPS
Consult GPS Specifications and Guidelines (NCCN 2006d, in development), GIS Naming Conventions (NCCN 2005a), and Design Standards for Project Databases (NCCN 2006b, in development) for GPS data collection specifications.

B. Tabular Data
Appendix 4 provides specific instructions for shapefile creation in ArcMap and ArcCatalog from dBase, text and Excel files using Microsoft Excel. Consult Design Standards for Project Databases (NCCN 2006b, in development) for tabular data standards.

C. On-Screen Digitizing And Scanning
The following section describes general work flow and recommended QA procedures for each step in GIS data set creation using on-screen digitizing and scanning collection methods. Not all steps are applicable to all collection methods. Applicable collection methods are indicated in parentheses after each section heading. Appendix 5 provides specific instructions for shapefile creation in ArcMap and ArcCatalog using on-screen digitizing. Appendix 6 contains instructions for shapefile creation in ArcScan using scanning.

i. Prepare paper map(s), mylar(s) or film (scanning)
Not all maps are ready to be scanned in their original state. Map scrub, or the preparation of maps, is the foundation for subsequent steps of the data conversion process. Detecting and correcting errors can be done most cost-effectively at this phase. What is required here is for the GIS Specialist to specify, in detail, a procedure for converting the map document into an acceptable digital file *while accounting for all known problems in the map document*. This procedure should be tested in the pilot project and modified as needed.

1) Control review
Benchmarks, corner tic marks, or other surveyed locations, visible and identifiable on the source map, are used to establish coordinate control for the new data set. Verify the known real-world location for every control point. This is the most important step in the data conversion process - every dollar spent on coordinate control is worth at least two dollars spent later dealing with positional accuracy problems.

2) Edgematch review
Edgematching requires that all features that cross or are near the map edge be reviewed with respect to logical and physical consistency requirements as well as evaluated for positional accuracy and duplication. The temporal factor must be considered. If adjacent maps differ greatly in age, there are bound to be edgematching errors between these maps. Cultural features are especially prone to this problem. Another confounding factor is that spatial errors on basemaps such as USGS quadrangles are typically concentrated in one corner or side of the map. The result is physical inconsistency between adjacent maps, which can in many cases be at least partially resolved after digitizing by using other base images, such as digital orthophotos.

3) Primary key validation
Map information that will be converted to primary key information must be closely reviewed. For example, if a fire hydrant number will be used as a primary key to relate to other tables in the database, the numbers must be checked so that all fire hydrant features have unique numbers to avoid referential integrity errors. Consult Design Standards for Project Databases (NCCN 2006b, in development) for NCCN primary key requirements.

4) Conflict resolution
Conflicts resulting when the same data coming from two or more sources differ must be worked out in order to select the best data to use for the project. Map series must be reviewed together to identify duplicated features and to resolve conflicting positional locations and conflicting feature attributes. Criteria for deciding between the sources are:
- Accuracy of resulting data
- Cost of conversion from source to database
- Availability of the source for conversion
- Availability of a continuing flow of data for database maintenance

Occasionally, alternative sources may result in different representations in the database, such as a vector representation versus a scanned image. In this situation, the ability of each representation to satisfy the requirements of the GIS applications will need to be evaluated.

5) Blemish removal
For scanned maps, blemishes on the source map may need to be removed from the resulting raster. For example, a fleck of dirt might connect two lines that should not be connected.

ii. Data conversion and revision (on-screen digitizing and scanning)

1) Spatial features
Digitize or scan features needed in the new data set. Instructions for creating a shapefile in ArcMap using on-screen digitizing are provided in Appendix 5. Implement the following QA procedures (for more specific methods see Johnston, 1995:

a) Positional accuracy

Implement visual QA procedures to check for positional accuracy problems. Visually inspecting data can detect systematic errors such as an overall shift in the data caused by an unusually high RMS value or random errors such as missing features. Visual QA can be performed using hard-copy plots or on-screen views. The hard-copy plotting of data is the best method for checking for missing features, misplaced features, and registration errors. On-screen views are an excellent way to verify that edits to the database were made correctly, but are not a substitute for inspecting plots.

Make sure any erratic or duplicate features have been edited out and alignment problems have been corrected. In cases where high positional accuracy needs to be confirmed or tested, the National Standard for Spatial Data Accuracy (NSSDA) (FGDC 1998) is a good place to start. The NSSDA describes a way to measure and report positional accuracy of features in a geographic data set. Consult GIS Product Specifications (NCCN 2005b) for NCCN I&M data set positional accuracy standards.

b) Physical consistency
Make sure that all features are topologically correct, i.e. all lines and polygons are properly connected and closed, respectively.

c) Projection and datum
Upon creation, projection and datum should be specified for all shapefiles (.pjr file) and coverages (use the *projectdefine* command in Arc). Add the exported or digitized data sets into a GIS program (like ESRI's ArcView or ArcMap) with other relevant GIS basedata to check for projection and datum problems. All of the data should align properly, within the scale and accuracy constraints of the data sets compared. If the digitized or scanned data set does not match basedata, then the collected data were probably in a different projection or datum. Reexamine the conversion methods to correct any data projection and datum problems.

2) Attribute information
Enter, if necessary, any attribute information collected on field forms or gathered from existing data sources.

a) Validity
Make sure field entries are standardized, complete, and accurate. Visual QA/QC can detect random errors such as misspelled text, presence of an erroneous date or the absence of data. Review and edit the attribute tabular data. Automated QA/QC procedures can be implemented to allow quick inspection of large amounts of data. Although automated QA/QC procedures may reveal random or systematic errors that were not detected during the visual inspection process, they should be used in addition to, rather than instead of, visual QA/QC procedures. Follow guidelines in the GIS Naming Conventions (NCCN 2005a) document when creating new attributes (fields).

b) Referential integrity

Make sure primary and foreign keys are present and accurate. No errors should be allowed in these fields, since incorrect information will result in errors cascading through the database and jeopardizing relationships to one or more tables. Perform any advanced data manipulation involving primary and foreign keys to check for errors. This could include any data combining with other data sets, creating hotlinks to photos, or performing a spatial join in order to trim features or add other attribute information. Consult the Spatial Data Integration (NCCN 2006g, in development) document for information on spatial relates and joins and the Design Standards for Project Databases (NCCN 2006b, in development) document for primary and foreign key requirements

c) Logical consistency
Verify that attributes that functionally relate to each other are entered accurately. Correct any errors.

4. Certify Data
After data collection and conversion processes are complete, the Project Lead or Principal Investigator should review the final data set for data completeness and compliance with accuracy standards and use predetermined criteria for acceptance of data set as accurate. Consult the Quality Assurance Certification (NCCN 2006f, in development) document for further information.

5. Document Data Set And Data Set Design

A. Data Dictionary
After the final data set has been reviewed and accepted, the Project Lead or Principal Investigator is responsible for updating the preliminary data dictionary to reflect any changes that might have occurred during the review process and, therefore, creating the final data dictionary.

B. Project Documentation
Adequate documentation throughout data development is vital for ensuring high quality metadata at the conclusion of a project. The Project Lead or Principal Investigator and GIS Specialist are jointly responsible for documenting their respective parts of the project. There is no standardized project documentation format, however, Metadata Interview form is a starting point (see below).

i. One way to track project process and progress is to keep a notebook of procedures and output from the project work and/or keep a digital folder to track all project work. The GIS Specialist's and Project Lead's or Principal Investigator's choice on how to track their work depends on the number of participants and complexity of the project. Documentation for a project can range from a simple Readme.txt file to a detailed daily log explaining processes and milestones.
ii. Documenting project objectives (Section 1A) and creating a preliminary data dictionary (Section 1J) will help the Project Lead or Principal Investigator to get started on the project documentation process.
iii. In addition, the Metadata Interview document (see Metadata Guidelines, (NCCN 2005e, in development)) can help both the GIS Specialist and the Project Lead or Principal Investigator collect the information needed for metadata creation.

C. Metadata

The Project Lead or Principal Investigator is responsible for compiling general information about the project, and for completing a Metadata Interview document for the project. Specific information about data collection methods and data processing should be compiled by the data collectors (field crew leaders) and/or data processors (GIS Specialists or Data Managers) for the project. This ensures minimal loss of important information. The Project Lead is responsible for compiling all of the metadata information into an appropriate metadata document. GIS Specialists (or Data Managers) are then responsible for creating a parsed metadata record from the materials provided by the Project Lead. See Metadata Guidelines, (NCCN 2006e, in development) for information of metadata creation and NPS and NCCN metadata requirements.

6. Post And Archive Data Set
GIS Specialists and Data Managers are responsible for posting complete and certified GIS data sets to NPS and NCCN data libraries and repositories. Data archival procedures should be followed as described in the Data Life Cycle section of the Data Management Plan (Boetsch et al. 2005) and in the File Storage and Archival Guidelines (NCCN 2006c, in development). Consult the Data Product Publishing and Distribution (NCCN, 2006a, in development) document and the Data Distribution section of the Data Management Plan (Boetsch et al. 2005) for details on the steps required for posting data and the types of data repositories.

7. Maintain Data Set
The data maintenance stage of the project life cycle begins once the final data set has been accepted by the Project Lead or Principal Investigator and reviewed by the GIS Specialist and the Data Manager. GIS data maintenance involves additions, deletions, and other updates to the data set and metadata. These changes must be done in a tightly-controlled manner in order to retain the database's integrity.

A. Check Out Procedures
If a certified and posted data set needs to be altered, the Project Lead or Principal Investigator should contact the GIS Specialist or Data Manager, who will check out the data set from permanent storage, copy it into local storage for updating, and then post it back to the permanent storage location. If more than one person makes changes to the data set and more than one version of the data set exists, the Project Lead or Principal Investigator should work with the GIS Specialist to reconcile the versions before the data set is posted back to permanent storage.

B. Data Alteration
Whenever the topology of a feature is altered (e.g. vertex deleted or added in a shapefile or in building/cleaning a coverage), attributes such as length and area must be recalculated. Unit values exported from the GPS field data may not match the final unit values in an edited GIS layer and should also be recalculated.

C. Data Merging
When combining GIS data sets, remember that the attributes tables usually need to match exactly in order for the attribute information from all input data sets to be included in the combined data set. The features usually combine properly, but one needs to look carefully at the attribute tabular data to ensure the fields are complete and standardized. See GIS Naming Conventions (NCCN 2005a) and Design Standards for Project Databases (NCCN 2006b, in development) for information on required attributes and standard field formats.

D. QA procedures

Before archiving and posting the data set following updates, QA checks are required to ensure data set integrity. The Data Manager and the GIS Specialist must maintain the data set schema so that table structure and spatial data topologies are not destroyed.

E. Metadata update
GIS Specialist and Data Manager are responsible for updating the metadata record for the data set following data alteration or data merging. Updated dataset is then posted and archived to appropriate NCCN and NPS repositories.

Responsibilities

- Project Leads (NPS), Principal Investigators (contractors and cooperators) and Data Managers working on NCCN I&M projects are responsible for consulting with NCCN GIS Specialists in all stages of GIS data development and, jointly with GIS Specialist, for providing training and support to project staff during I&M GIS data development within the Network.
- The Project Lead or Principal Investigator and GIS Specialist will agree on spatial data processing responsibilities before full scale data collection begins.
- NCCN GIS Specialists and Data Managers are responsible for developing GPS data dictionaries, appropriate QA procedures and acceptance standards for new GIS data sets.
- Project Lead or Principal Investigator will verify that all newly-created GIS data sets meet the requirements in these documents prior to data set submission to GIS Specialist and/or Data Manager for posting or archiving.

Credits

- Colonial National Historic Park. 1998. SOP for Themes, Attribute Tables, and Map Design for Colonial NHP, NPS. USDI National Park Service. Available at: http://nps.gov/gis/data_info/documents/colo_gis_sop_95.pdf
- ESRI. 2005. Data Model. Available at: http://www.esri.com/software/arcgis/geodatabase/about/data-models.html
- ESRI. 2005. The Geodatabase View. Available at: http://www.esri.com/software/arcgis/concepts/gis-data.html
- ESRI. 2003. ArcGIS 8.3 Desktop Help Files.
- Geographic Resource Information Management – Intermountain GIS. 2005. Technical Tips: howto files and helpful scripts. USDI National Park Service. Available at: http://imgis.nps.gov/technical_tips.html
- NCGIA. 1998. GIS Core Curriculum for Technical Programs. National Center for Geographic Information and Analysis (NCGIA). Available at: http://www.ncgia.ucsb.edu/cctp/
- New York State Archives. 1996. Geographic Information System Development Guides. Available at: http://www.nysarchives.org/a/nysaservices/ns_mgr_active_gisguides.shtml
- Probala, Jeanna. 2005. Surficial Geology Mapping Protocol – Draft. USDI National Park Service, North Coast and Cascades Network.
- UNESCO. 1999. Training Modules on the Applications of Geographic Information Systems (GIS). Available at: http://ioc.unesco.org/oceanteacher/resourcekit/Module2/GIS/Module/index.html
- University of Melbourne. 2005. GIS Self Learning Tool. University of Melbourne, Department of Geomatics. Available at: http://www.sli.unimelb.edu.au/gisweb/SDEModule/SDE_Theory.htm

Recommended Citation

North Coast and Cascades Network – National Park Service. 2005. GIS Data Layer Design and Creation Guidelines. USDI National Park Service.

Contact Information

Mount Rainier National Park Tahoma Woods – Star Route Ashford, WA 98304	Darin Swinney, GIS Coordinator 360-569-2211 ext. 3378 darin_swinney@nps.gov
North Cascades National Park Service Complex 810 State Route 20 Sedro-Woolley, WA 98284-1239	Anne Braaten, GIS Coordinator 360-856-5700 ext. 238 anne_braaten@nps.gov
Olympic National Park 600 East Park Avenue Port Angeles, WA 98362-6757	Roger Hoffman, GIS Coordinator 360-565-3062 roger_hoffman@nps.gov
Ebey's Landing National Historical Reserve	Please refer to contacts for North Cascades NP
Fort Vancouver National Historic Site	Please refer to contacts for Mount Rainier NP
Lewis and Clark National Historical Park	Please refer to contacts for Mount Rainier NP

San Juan Island National Historical Park	Please refer to contacts for North Cascades NP

Reference Documents

Related Guidance

- Boetsch, J.R., B. Christoe, and R.E. Holmes. 2005. Data management plan for the North Coast and Cascades Network Inventory and Monitoring Program. USDI National Park Service. Port Angeles, WA. 88 pp.
- Natural Resource Database Template Version 3 Documentation. 2004. Washington Support Office, Inventory and Monitoring Program. National Park Service. Fort Collins, CO. Available at: http://science.nature.nps.gov/im/apps/template/index.htm.
- National Park Service GIS Program. 2005. NPS GIS Data Specifications for Resource Mapping, Inventories and Studies. USDI National Park Service. Available at: http://science.nature.nps.gov/nrgis/standards/docs/GISSpecs31105_final.pdf.
- North Coast and Cascades Network – National Park Service. 2006a. Data Product Publishing and Distribution. USDI National Park Service. Available at: http://www1.nature.nps.gov/im/units/nccn/dm_docs.htm (in development).
- North Coast and Cascades Network – National Park Service. 2006b. Design Standards for Project Databases. USDI National Park Service. Available at: http://www1.nature.nps.gov/im/units/nccn/dm_docs.htm (in development).
- North Coast and Cascades Network – National Park Service. 2006c. File Storage and Archival Guidelines. USDI National Park Service. Available at: http://www1.nature.nps.gov/im/units/nccn/dm_docs.htm (in development).
- North Coast and Cascades Network – National Park Service. 2005a. GIS Naming Conventions. USDI National Park Service. Available at: http://www1.nature.nps.gov/im/units/nccn/dm_docs.htm.
- North Coast and Cascades Network – National Park Service. 2005b. GIS Product Specifications. USDI National Park Service. Available at: http://www1.nature.nps.gov/im/units/nccn/dm_docs.htm.
- North Coast and Cascades Network – National Park Service. 2006d. GPS Specifications and Guidelines. USDI National Park Service. Available at: http://www1.nature.nps.gov/im/units/nccn/dm_docs.htm (in development).
- North Coast and Cascades Network – National Park Service. 2006e. Metadata Guidelines. USDI National Park Service. Available at: http://www1.nature.nps.gov/im/units/nccn/dm_docs.htm (in development).
- North Coast and Cascades Network – National Park Service. 2006f. Quality Assurance Certification. USDI National Park Service. Available at: http://www1.nature.nps.gov/im/units/nccn/dm_docs.htm (in development).
- North Coast and Cascades Network – National Park Service. 2006g. Spatial Data Integration. USDI National Park Service. Available at: http://www1.nature.nps.gov/im/units/nccn/dm_docs.htm (in development).

Other Citations and References

- FGDC. 1998. Geospatial Positioning Accuracy Standards, Part 3: National Standard for Spatial Data Accuracy, FGDC-STD-007.3-1998. Available at: http://www.fgdc.gov/standards/status/sub1_3.html
- Ghose, M.K. Unknown date. Product quality assurance for GIS life-cycle. ISRO, Dept. of Space, Kharagpur. Available at: http://www.gisdevelopment.net/technology/gis/techgi0042.htm

- Johnston, D.M., D.M. Timlin, D.L. Szafoni and J.J. Casanova. 1999. Quality Assurance/ Quality Control Procedures for ITAM GIS Databases. Geographic Modeling Systems Lab, University of Illinois at Urbana, Champaign. Available at: http://www.gis.uiuc.edu/research/spatialanalysis/Quality%20Assurance.htm
- McCain, M.J. and W.C. Masters. 1998. Integrating Quality Assurance into the GIS project life cycle. ESRI ArcUser Magazine. Available at: http://www.esri.com/news/arcuser/798/qa.html
- Lillesand, T.M. and R.W. Kiefer. 1994. Remote Sensing and Image Interpretation. John Wiley & Sons, Inc., New York.
- NPS GIS Program. 2003. GPS for GIS workflow. USDI National Park Service. Available at: http://www.nps.gov/gis/gps/gps4gis/
- US Bureau of the Budget. 1947. National Map Accuracy Standards. National Mapping Program Standards, USGS. Available at: http://rockyweb.cr.usgs.gov/nmpstds/nmas.html

Revision History

Revision Date	Description of Change	Author	Effective Date
Mmm dd, yyyy		Full name	Mmm dd, yyyy

Appendix 1. Spatial analysis with GIS.

- Data (representation) modeling
  Data modeling can involve both vector and raster layers.
    i. Vector representation models try to describe the features in a landscape such as buildings, streams or forests using a formalized, conceptual schema, which is usually implemented using points, lines and polygons. The representation model attempts to capture the spatial characteristics of a feature (the shape of a building) and relationships among features in the landscape (the distribution of buildings). Along with establishing the spatial relationships, the GIS representation model is also able to model the attributes of the features (who owns each building). Vector layers can also estimate characteristics of surfaces from a limited number of point measurements. For example, a GIS can quickly generate lines that indicate rainfall amounts, along isoclines, using data from individual weather stations. The resulting GIS data can be thought of as a rainfall contour layer. This two-dimensional rainfall contour layer may then be overlaid and analyzed with other GIS data covering the same area.
    ii. Raster data are generally divided into two categories, thematic data and image data. The values in thematic raster data represent some measured quantity or classification of a particular continuous phenomena, such as elevation, pollution concentration, or population. For example, in a landcover map the value 5 may represent forest and the value 7 may represent water. On the other hand, the values of cells within image data represent reflected or emitted light or energy, such as that of a satellite image or scanned photograph.

- Topological modeling
  A GIS can be used to identify and analyze the spatial relationships that exist within digitally stored spatial data. Topological relationships between geometric entities traditionally include adjacency (what adjoins what), containment (what encloses what), and proximity (how close something is to something else). For example, are there any gas stations or factories operating next to the swamp (adjacency)? Are there any within two miles and uphill from the swamp (proximity)? Such topological relationships allow complex spatial modeling and analysis to be performed.

- Networks
  A GIS can simulate the routing of materials along a linear network. Values such as slope, speed limit and pipe diameter can be incorporated into network modeling in order to represent the flow of the phenomenon more accurately. Network modeling is commonly employed in transportation planning, hydrology modeling, and infrastructure modeling. For example, if all the factories near a wetland were accidentally to release chemicals into the river at the same time, how long would it take for a damaging amount of pollutant to enter the wetland reserve?

- Cartographic (process) modeling
  Cartographic modeling involves using cartographic form (i.e. maps) to depict the interaction of objects synthesized in a representation model. It is a technique used for both vector and raster based GIS, and it is used to simultaneously analyze both the spatial and thematic characteristics of geospatial information. The thematic component of geospatial information is analyzed via statistical operations on the data (for example, calculating the average and standard deviation of the data), while the spatial characteristics of geospatial information are described through spatial

analysis techniques (which are based on coordinate data). Cartographic models are most often used to predict what will happen if some action occurs.
There are many types of cartographic models including:

i. Suitability modeling: Most spatial models involve finding optimum locations such as finding areas suitable for lynx habitat or areas suitable for logging.
ii. Distance modeling: What is the walking distance between the trailhead and the summit?
iii. Hydrologic modeling: Where will the water flow to?
iv. Surface modeling: What is the geographic extent of the watershed? What is the pollution level for various locations in a watershed?

- Vector overlay
  Vector overlay is the combination of two separate spatial data sets (points, lines or polygons) to create a new output vector data set. Common vector overlay procedures include Union, Intersection, and Identity operations. All three overlay operations compute the geometric intersection of two vector data sets. Union overlay combines all the geographic features of both input vector data sets. Intersection overlay preserves only those features in the area common to both data sets. Identity overlay preserves all features of the input data set, as well as those features of the identity data set that overlap the input data set. The fields of the attribute tables of the two input vector data sets are combined into the attribute table of the new, output data set.

- Spatial statistics
  Spatial statistics is the collection of statistical methods in which spatial locations play an explicit role in the analysis of data. Spatial data in most cases are not spatially independent. Data values which are close spatially show less variability than data values which are farther away from each other. For example, pollution tends to be concentrated closer to the source and/or natural barriers to dissipation. Spatial statistics can analyze patterns (measure spatial autocorrelation concentration and nearest neighbor distance), map clusters and measure geographic distribution (determine the location of the center of the data, the shape and orientation of the data, and the degree to which features are dispersed). In addition, geostatistics provide means for generating continuous surfaces from selected data points.

- Geocoding
  Geocoding refers to calculating spatial locations (x,y coordinates) from events, which are usually provided in the form of a table or database. A reference theme is required to geocode individual events, such as a stream centerline file with distance ranges. The individual event locations are interpolated, or estimated, by examining distance ranges along a stream segment. For example, in a National Hydrography Dataset stream layer a event point of 500 will be at the midpoint of a specified stream reach that starts with distance 0 and ends with distance 1000. The GIS will then place a dot approximately where that event belongs along the segment of centerline.

- Reverse geocoding
  Reverse geocoding refers to estimating an event number from a given set of coordinates. For example, a user can mouse-click on a stream centerline theme (thus providing a set of coordinates), and the GIS will estimate the distance along the stream reach. This distance is interpolated from a range assigned to that stream segment. If the user mouse-clicks at the midpoint of a segment with distances ranging from 1 to 100, the returned value will be approximately 50.

Appendix 2. Evaluation criteria for a data source.

Map series, photos, remotely sensed images, and tabular files are typical sources of data for creating new GIS data sets. Before they can be used as such, they must be reviewed and evaluated for suitability to use in the GIS. Critical elements of these data sources that must be evaluated include (note that not all apply to tabular data):

- Appropriate scale
- Projection and coordinate system
- Availability of geodetic control points
- Aerial coverage
- Completeness and consistency across entire area
- Symbolization of entities (especially positional accuracy of symbol due either to size of symbol or off-set placement on map)
- Quality of linework and symbols
- General legibility for digitizing (labels)
- Quality and stability of source material (paper/mylar)
- Amount of manual editing needed prior to conversion
- Edgematch between map sheets
- Existence and type of unique identifiers for each feature - often features shown on map series use so-called "intelligent" keys or identifiers where an identifier for an object contains the map sheet number and/or other imbedded locational codes. For example, a wetland’s unique identifier might contain “H348121” (a USGS map sheet identifier) or “CM” (an abbreviation for USGS Crater Mountain quad). In database design, it is much better to avoid "intelligent" keys of this type.
- Positional and attribute accuracy
- Data source date
- Presence of metadata

Appendix 3. GIS data collection methods

- GPS
  With GPS, positions are collected directly in a digital form in the field, and both spatial features (points, lines, polygons) and attributes are collected (if a data dictionary is loaded simultaneously, thereby eliminating transcription errors). Consult GPS Specifications and Guidelines (NCCN 2006d, in development) for NCCN requirements on GPS data collection. Keep in mind that different GPS units collect data with different positional accuracies, and not all GPS units support data dictionaries, which are necessary to record attribute data electronically. These considerations should be addressed when selecting a GPS unit for a project. In addition, field crew leaders need to make sure that the GPS unit is properly configured. Project Leads or Principal Investigators (with help from GIS Specialist) need to make sure that a data dictionary is carefully designed before field data collection begins.

- Digitizing from paper maps
  Digitizing is the process of converting source materials, prepared manually, into the digital vector format stored and processed by computers. Digitizing involves tracing map features one at a time into a computer using a digitizing tablet, graphics tablet, mouse, or keyboard-controlled cursor. With the advance of scanning capabilities, NCCN no longer recommends digitizing from paper maps.

- Scanning from paper maps
  Scanning is a process by which data and maps are converted to digital form using optical or video equipment. Scanning differs from digitizing in that entire pages of data or map sheets are captured as images at once in raster format, which then can be converted to vector format. Once in a computer, scanned images require coding; the operator must indicate the data to be stored, how it is attributed, and where to store the information.

- On-screen digitizing
  On-screen digitizing is an interactive process in which a data set is created using previously digitized or scanned information (such as DOQs) as a reference. This method of digitizing is commonly called "heads-up" digitizing because the attention of the user is focused up on the computer screen and not on a digitizing tablet. This technique may be used to trace features from a scanned map or image to create new GIS data sets.

- Remote sensing
  Remote sensing is the science and art of obtaining information about an object, area or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area or phenomenon under investigation (Lillesand and Kiefer 1994). Remote sensing is accomplished with either passive sensors measuring the reflectance from the earth’s surface in certain regions of the electromagnetic spectrum, or active sensors measuring radio waves that were sent out from the sensors themselves. Remote sensing technologies collect raster data that can be further processed to identify objects and classes of interest, such as land cover.

- Photogrammetry
  Soft copy workstations (composed of an image high-resolution scanner, high-speed computer processor, large high-resolution monitor, and appropriate software for viewing scanned images in 3-D, drawing and editing planimetric and topographic information, and translating data sets to various formats for the end user) are used to digitize features directly from stereo pairs of digital photographs. These systems allow data to be captured in 2 and 3 dimensions, with elevations

measured directly from a stereo pair using principles of photogrammetry. Currently, analog aerial photos are scanned before being entered into a soft copy system, but as high quality digital cameras become cheaper this step will be skipped.

Appendix 4. ArcMap: Creating new shapefiles from dBase/text/Excel files using MS Excel spreadsheets.

(Adapted from the NPS, Intermountain Geographic Resource Information Management Team, Denver. 2/15/2005. Modified for NCCN on 1/19/2006)

These instructions address how to convert a tabular list of point coordinates into a new ESRI shapefile. The users can also use them to save data to a geodatabase feature class.

1. Start MS Excel. Open the file with the data and coordinate information by selecting File → Open. Navigate to file location and change file type to "Text Files" or "dBase Files," if necessary. Select file and click on "Open." If using a text file, navigate through Text Import Wizard. Field headings must at least include: GIS_Loc_ID, Northing, Easting or GIS_Loc_ID, Latitude, Longitude. See Design Standards for Project Databases (NCCN 2006b, in development) for specifications on GIS_Loc_ID. See GIS Naming Conventions (NCCN 2005a) for guidelines on format and naming of fields other than GIS_Loc_ID. If using geographic coordinates, the format of the coordinates must be in decimal degrees and a negative sign for west Longitude. For geographic coordinates, it is best to have at least 5 or more significant digits after the decimal to make sure the data are precise enough for large scale mapping.

    A. To convert degrees, minutes, seconds to decimal degrees use the following formula:
    DMS → dd: D + M/60 + S/3600 = dd
    Example: 45°30'45.22"→ 45 + 30/60 + 45.22/3600 = 45.512561

2. Clean up the file. Make any hidden columns visible, then delete those that will not be needed in the final GIS shapefile. Remove all empty rows between the header row and data. Empty entries in the coordinate columns will not be brought over.

3. Make sure the Northing and Easting (or Latitude and Longitude) columns are formatted to 'number'. If there are no decimal places, specify that in the format dialog box (select column → Format dropdown menu → Cells → Number → Decimal Places). Select all columns, then select Format → Column → Autofit Selection.

4. Save the file as an Excel file now to serve as a backup.

5. Put the curser in the cell directly below the last lower right cell holding the data or select all cells containing data.

6. Save the file to a format that ArcMap can read using the following steps:

    A. File menu → Save As.

    B. In the Save As Dialog box, Select as type: CSV (Comma delimited,*.csv) or dBase IV (*.dbf).

    C. Name the file appropriately and press Save. File name should not exceed 8 characters.

    D. Select "Yes" to the warning that Excel does not accept workbooks that contain multiple sheets.

    E. Select "Yes" to the warning that incompatible formats exist.

F. Close the File and select NO to the next warning to save changes.

7. Open an existing ArcMap document or create a new one. It is important to have the data frame projection defined. Do this by double clicking on the data frame's name (usually 'Layers') in the TOC portion of ArcMap. Layers → Coordinate System tab

8. Start the dialog box to add XY data by selecting Tools → Add XY Data.

9. The Add XY tool allows showing the location of XY coordinates from a .dbf or .csv file on-the-fly. Fill in the following sections of the dialog box and then click OK.

    A. Name (navigate to the .csv or .dbf file using the folder button on the right).

    B. Choose the correct X and Y field headings from the table. If the field headings were labeled as Northing/Easting or Latitude/Longitude, the Add XY dialog box automatically fills in the fields for the X and Y coordinates.

    C. Set the Spatial Reference (click on the "Edit" button). The program needs to know what projection and datum the coordinates are in.

10. The data will display in the data frame map area as an event layer (not yet a shapefile!). Check to make sure all of the positions are present and that there are no errors in the attributes, if any, or relative locations.

11. Save the data to a shapefile. Right click on the filename in the Table Of Contents (TOC, the left pane)→ Data → Export Data → Shapefile. Be sure to read the Export Data dialog box carefully for information on the coordinate system to assign to the data. Consult GIS Naming Conventions (NCCN 2005a) for guidelines on naming new GIS files.

Further information can be found in:
+Using_ArcMap.pdf (digital book)
+ArcGIS Desktop Help → Contents tab → ArcMap\Creating maps\Adding x,y coordinate data to a map
+ArcGIS Desktop Help → Contents tab → ArcCatalog\Exploring the values in a table\About tabular data sources

Appendix 5. ArcGIS: Creating new shapefiles using on-screen digitizing

(Adapted from the NPS, Intermountain Geographic Resource Information Management Team, Denver. 9/22/2004. Modified for NCCN on 1/19/2006)

These instructions address making a new shapefile data set and populating it by on-screen digitizing features. Both ArcCatalog and ArcMap are utilized. The instructions can be adapted to apply to a geodatabase feature class.

1. Open ArcCatalog.

2. In ArcCatalog, navigate to the location where the new data will be created. This will probably be in a "working" directory on a local machine or in a special folder on a GIS data server.

3. Create a new data set. Right click in the Contents Tab (on the right side panel) or right click on the directory name (left side panel) and choose New → Shapefile. A dialog box appears. Specify the following:

   A. Name of the data set. See GIS Naming Conventions (NCCN 2005a) for guidelines on GIS layer names.
   B. Feature Type (point, polyline, or polygon)
   C. Spatial Reference (hit the "Edit" button and in the new dialog box use the "Select" button to find: UTM, zone 10N, meters, NAD83 (NPS and NCCN standard projection))

4. Open ArcMap. Start a new project and save it or open an existing project.

5. Check the data frame's coordinate system. It needs to be set to the coordinate system that the background and new data set are in. Do this by double clicking on the data frame's name (usually 'Layers') in the TOC portion of ArcMap. Layers → Coordinate System tab.

6. Add any background data needed for relative placement of the new on-screen digitized data. Use the "Add Data" button or drag and drop data from ArcCatalog. Most of the time this will be raster data such as photos or DRGs. Make sure the background data are in the same coordinate system and datum as the newly created shapefile.

7. Add the new, blank shapefile created in Step 3.

8. To add fields to the attribute table before starting to digitizing features, right-click on the data set's name in the TOC → Open Attribute Table to view the attributes. Use the "Option" button to Add Fields. See GIS Naming Conventions (NCCN 2005a) for guidelines on field names and Design Standards for Project Databases (NCCN 2006b, in development) for information on NCCN required fields.

9. Make sure the *Editor* Toolbar is displayed. Do this by going to the View Menu → Toolbars → Editor. Once it is on, it can be dragged and "docked" somewhere.

10. Start Editing the data by going to *Editor* Toolbar → Editor Menu → Start Editing. Depending on what data are loaded into ArcMap, a dialog box may appear asking which directory holds the data to be edited. Be sure to select the one with the new, blank data set.

11. Take a look at the *Editor* Toolbar's TASK and TARGET dropdown boxes. For purposes of creating new data, these need to be set to: TASK → Create New Data and TARGET → name of new data set.

12. Zoom to the area of interest where features will be digitized into the new data set.

13. Click on the pencil icon that is part of the *Editor* Toolbar. This is the simplest sketch tool.

14. Features can now be traced or drawn on the data frame's map display. Single-click to create points and single-click to start a line or polygon. To end a line or polygon, double-click on the last vertex or right-click and select "Finish Sketch." Note: Save the edits often by going to *Editor* Toolbar → Editor Menu → Save Edits.

15. To view or enter attribute information as features are created, right-click on the data set's name in the TOC → Open Attribute Table. At this point, the newly created feature will be selected (highlighted) in the table. Enter attributes by right-clicking in one of the cells. Alternatively, fill in attributes after digitizing all the features (see #19).

16. To modify a feature, change the TASK to "modify feature" click on the small black arrow tool on the *Editor* Toolbar, and click on the feature to be modified to select it.

17. To delete a feature, select it with the small black arrow tool on the *Editor* Toolbar and then hit the keyboard's delete key.

18. When done adding and modifying data, go to the *Editor* Toolbar → Editor Menu → Save Edits & Stop Editing.

19. To edit or add data to the shapefile's tabular data, right-click on the data set's name in the TOC → Open Attribute Table. Note: Use the "Option" button to Add Fields when the data set is not in editing mode. The new fields can be populated once editing is resumed (see Step 10).

Appendix 6. ArcGIS: Creating new shapefiles using ArcScan.

These instructions address conversion of raster images into vector-based shapefiles using ArcScan.

Notes:

- Scanning is an efficient way for a large amount of legacy or field-generated geographic information, in the form of hard-copy maps, to be integrated into a GIS, given the source data can be readily reclassified into two values.
- Currently, ArcScan only works on bitonal (1-bit) image classifications; best results come from media that are already bitonal. Sources, however, should be scanned as 8-bit grayscale images with at least 300 dots per inch (dpi). Grayscale images can then be reclassified in ArcMap to black and white before vectorization. Reclassifying in ArcMap provides the best results with regard to linework because the reclassification can be executed within an ArcScan session, with changes to vectors being viewable on-the-fly.
- Regardless of the quality of the source data and of the resulting scanned image, some editing of raster and/or vector data will be necessary. The most efficient process involves altering the raster before vectors are generated. This approach yields vectors that require little to no editing.
- It is highly recommended that the user read all relevant ArcMap help files for more background information and detailed instructions of each step before starting the scanning and vectorization process.

1. Create a new point, line or polygon shapefile in ArcCatalog using Steps 1-3 in Appendix 6.

2. Using a large format scanner, convert each paper or mylar map into a digital raster-based image. Scan source to an 8-bit (64 value) grayscale image (TIFF) at 300 dpi (minimum).

   A. To determine the needed resolution, identify the smallest object to be included from the image. The pixel size must be equal in width to that smallest object. A pixel can only have one value, and that value will be an average of all values collected within that area. For a single object to be the determining factor for the value of a pixel it must cover at least half of the area of that pixel. For example, if a line half as wide as a pixel straddles two pixels, only 1/4 of each pixel will be covered by the line. The line will not be recorded.

      Scanning determines pixel size by dpi. If an image is scanned at 150 dpi, a linear inch has been sampled 150 times. To effectively scan at 150 dpi, any feature to be included must be at least 1/150 of an inch wide. It is important to choose a resolution that will effectively convey the information on the map but not to such a high degree it becomes hard to manage or exceeds file size limitations. For example, an image scanned at 150 dpi and using 5 Mb of storage space will require 334 Mb of storage if scanned at 1200 dpi.

   B. The source medium should not rotate or slip during the scan process.

   C. Make sure to include at least four control points or registration marks the coordinates of which are known on the scanned image. These can be points where UTM or Latitude/Longitude lines intersect on a 7.5' USGS quad (corners of map are best) or points whose locations were recorded in the field using GPS.

3. Resample image if needed.

4. Set coordinate system for the scanned image.
   A. In ArcCatalog, select the scanned image for which the coordinate system needs to be defined.

B. Click the File menu, and click Properties.
C. Click the Spatial Reference tab.
D. Click Edit.
E. In the Spatial Reference Properties dialog box, click Select to choose a predefined coordinate system. This coordinate system should match the coordinate system of the original map or mylar.
F. Click OK in the Raster Properties dialog box.

5. Create an x,y table in MS Excel populated with the known coordinates of at least four control points or registration marks from the source map. Using the procedures in Appendix 4, convert the table into a shapefile in ArcMap to be used as target layer for georeferencing.

6. Add scanned image to ArcMap and build pyramids when prompted.

7. Georeference the scanned image.

   A. In the Table of Contents, right-click the target layer (shapefile with control point coordinates) and click Zoom to Layer.
   B. From the Georeferencing toolbar, click the Layer dropdown arrow and click the raster layer (scanned image) to georeference. If the Georeferencing toolbar is not showing, click on View menu → Toolbars → Georeferencing.
   C. Click Georeferencing and click Fit to Display. This will display the raster in the same area as the target layers. Use the Shift and Rotate tools to move the raster as needed.
   D. Click the Control Points button to add control points.
   E. To add a link, click the mouse pointer over a known location on the raster, then over a known location on the target data. It may be useful to use a magnification window to add the links in.
   F. Add enough links for the transformation order. A minimum of three links is needed for a first-order transformation, six links for a second-order, and 10 links for a third-order.
   G. Click View Link Table to evaluate the transformation. Examine the residual error for each link and the RMS error. If the registration is satisfactory, stop entering links. Acceptable RMS values will vary depending on the accuracy of the original data and the scale of the source map. Determine an acceptable RMS value by examining the data and the scale it represents. A perfect transformation produces an RMS error of 0.000. High RMS errors indicate that the control points of the source map and the new control points of the layer that is being georeferenced do not correspond to the same relative locations. This can happen for one of the following reasons: original coverage registration marks (tics) were not digitized accurately, there is map sheet distortion such as stretching or shrinking, or the real-world coordinate values for tics were not recorded correctly (for example, an x- and y-coordinate value might be flipped for one of the tics).
   H. Click Georeferencing and click Update Georeferencing to save the transformation information with the raster. This creates a new file with the same name as the raster but with an .aux file extension.

8. Make sure statistics can be calculated for the georeferenced image(s). In ArcCatalog, right-click on the image and click Calculate Statistics. If the image lacks calculation then go through the Calculate Statistics dialog box that appears by putting a number 1 in the Row and Column boxes. Click OK.

9. Mosaic georeferenced images if there are more than one and they are spatially adjacent.

A. In ArcMap click the Add Data button on the top of the tool bar and navigate to the directory containing the scanned images. Add all spatially referenced images that will be combined (make sure each image has pyramids and statistics were calculated). Up to 50 images can be combined at one time.
B. Represent all rasters as bitonal images. In the TOC, right click on each raster and click on Properties. Select Symbology tab. In the Show drop down box, select Unique Values and click OK.
C. Double-click on the new white symbology box, select Properties tab and select "the color is null." Click OK. The goal is to have the image displayed in two colors – black for lines to be turned into vector data and white for background.
D. If after Step C the rasters display with more than 2 colors or a strip of no data exists where rasters overlap, click Classified under the Symbology tab and select 2 classes. Change symbol of linework to black and background to white.
E. Click on the Tools menu → Extensions → Spatial Analyst. If the Spatial Analyst toolbar does not appear, click on the View menu → Toolbars → Spatial Analyst.
F. On the Spatial Analyst toolbar click "Spatial Analyst" and scroll to and click "Raster Calculator." In the Raster Calculator dialog box type *mosaic*, add a left parenthesis, from the Layer box double-click on each image to be combined with a comma and space between each, and add a right parenthesis. For example: *Mosaic(.tif, .tif, tif)*. Click Evaluate. Depending on how large the images are it may take awhile. Once the calculation is complete a new layer will appear in the TOC. If the equation is not accepted, try any of the following:
   i. Change working directory. Make sure the working directory folder name is no longer than 8 characters with no spaces or symbols. Click on the Spatial Analyst icon → Options → General tab → navigate to the working directory → OK. It is recommended with large files to keep the scanned images and calculations in a temporary folder. Once the images are final, they can be saved to a permanent file in another directory.
   ii. Change Extent. Click on the Spatial Analyst icon → Options → Extent tab → Analysis Extent drop down box → Union of Inputs → OK.
   iii. Check and change spatial reference. Make sure that the coordinate system selected for the scanned images has larger extent than the new layer/image.
G. Remove all the TIFF images from the map and keep the Calculation layer.

10. Raster Cleanup
Before the digital images are converted into points, lines or polygons, they will require some preparation. This involves removing noise and raster elements that should not be vectorized. It also involves adding new features or filling gaps in the data, which will enhance the success of the vectorization.

A. Click the "Add Data" icon and navigate to the shapefile created in Step 1. Add it to the View. It may help to change the default colors to black and white in the layer's properties.

B. Start Editing by clicking the Editor Toolbar → Start Editing. If necessary, change the Task box to Create New Features and the Target box to Line. Change the Raster box on the ArcScan Toolbar to Calculation. If the Editor Toolbar is not present, click on the Tools menu → Editor Toolbar. If the ArcScan Toolbar is not present, click on the Tools menu → Extensions → ArcScan.

C. On the ArcScan Toolbar click Vectorization → Show Preview. Red vector lines will appear on the screen tracing over the raster lines. These lines are temporary and should be used as a guide while image cleanup is in progress. Adjust tolerances for creating the vector lines by using Vectorization → Vectorization Settings. Suggestion: Select the polygon style under Styles to start.

D. Zoom in on an area so that only a few lines or polygons are in the View. Every line segment will need to be reviewed closely, especially near the image boundaries. Some lines will appear in blue indicating problem areas identified by ArcScan. Small erroneous polygons within the linework and “noise,” which usually floats between linework, will need to be deleted.

E. Click on the ArcScan Toolbar → Raster Cleanup → Start Cleanup. Select Raster Cleanup → Raster Toolbar. Several cleanup tools will appear. With Preview on, use raster editing tools to erase or add raster ‘cells’ to smooth vector lines. Systematically pan through the data set, editing the raster data as needed. The Swap icon allows erasing or drawing with various box sizes. Use the magic eraser if there is a lot of “noise” on the image. Special attention should be given to erasing the tic marks, connecting lines from image to image, erasing small bubbles within a line, and erasing noise.

F. When the editing is complete, select Raster Cleanup → Stop Cleanup. Select Editor → Stop Editing → Save Edits.

11. Vector Generation

A. Start editing and from the ArcScan toolbar (??) select Vectorization → Generate Features. Select Line in the line layer box and/or Polygon in the polygon layer box. Click OK. Save the resulting vector data by clicking Editor → Save Edits. Close ArcMap.